

SaulLM-7B: A pioneering Large Language Model for Law

Year: 2024 | Citations: 112 | Authors: P. Colombo, T. Pires, Malik Boudiaf

Abstract

In this paper, we introduce SaulLM-7B, a large language model (LLM) tailored for the legal domain. With 7 billion parameters, SaulLM-7B is the first LLM designed explicitly for legal text comprehension and generation. Leveraging the Mistral 7B architecture as its foundation, SaulLM-7B is trained on an English legal corpus of over 30 billion tokens. SaulLM-7B exhibits state-of-the-art proficiency in understanding and processing legal documents. Additionally, we present a novel instructional fine-tuning method that leverages legal datasets to further enhance SaulLM-7B's performance in legal tasks. SaulLM-7B is released under the MIT License.